EGYPTIAN SPACE AGENCY

EGYPTIAN UNIVERSITIES TRAINING SATELLITE PROJECT

EUTS

ANOMALY DETECTION FOR SATELLITE TELEMETRY DATA USING MACHINE LEARNING

Prepared by

Helwan University

Computer and Artificial intelligence–Computer Science T42

Egyptian Universities Training Satellite

Introduction

- **Project Idea**: detect of telemetry data that come from satellites with using of machine learning algorithms,
- **Telemetry data** is set of measurement and reading of embedded device Across time interval
- **In this project** ,we use two dataset consist of Date and Temperature records ,after preprocessing dataset we use time series and machine learning algorithm to detect future temp

EgSA
وكالة الفضاء المصرية
Egyptian Space Agency

Data Analysis

EUTS

Egyptian Universities Training Satellite

Dataset 1

- First we load dataset 1 and 2 and try to figure them out, by see column , datatypes…etc.
- dataset 1 consists of one column "Date","Temp" and we try to separate them into two columns ‘Date’ , ‘temp’
- Convert temp column data type from string to float

```
"1986-02-04"" at position 1859
```

- Convert Date column data type from string to Datetime
- Drop Null values and check for Duplicated
- Finally data after Cleaned

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3647 entries, 0 to 3649
Data columns (total 2 columns):
Date     3647 non-null datetime64[ns]
Temp     3647 non-null float64
dtypes: datetime64[ns](1), float64(1)
memory usage: 85.5 KB
```

Dataset 1

```
"1986-02-04"" at position 1859
```

- First we load dataset 1 and 2 and try to figure them out, by see column , datatypes…etc.
- dataset 1 consists of one column "Date","Temp" and we try to separate them into two columns ‘Date’ , ‘temp’
- Convert temp column data type from string to float
- Convert Date column data type from string to Datetime

Dataset 2

- First we load dataset 2 and try to figure them out, by see column , datatypes…etc.
- dataset 2 it consists of one column "Month","Sunspots" "Date","Temp" we try to seprate this column into two column Date and Sunspots
- convert data type for Sunspots from String to float
- Convert Date column data type from string to Datetime
  - we found two issue at index 7 where year and month are swapped ,at index 489 invalid month
- Drop Null values and check for Duplicated

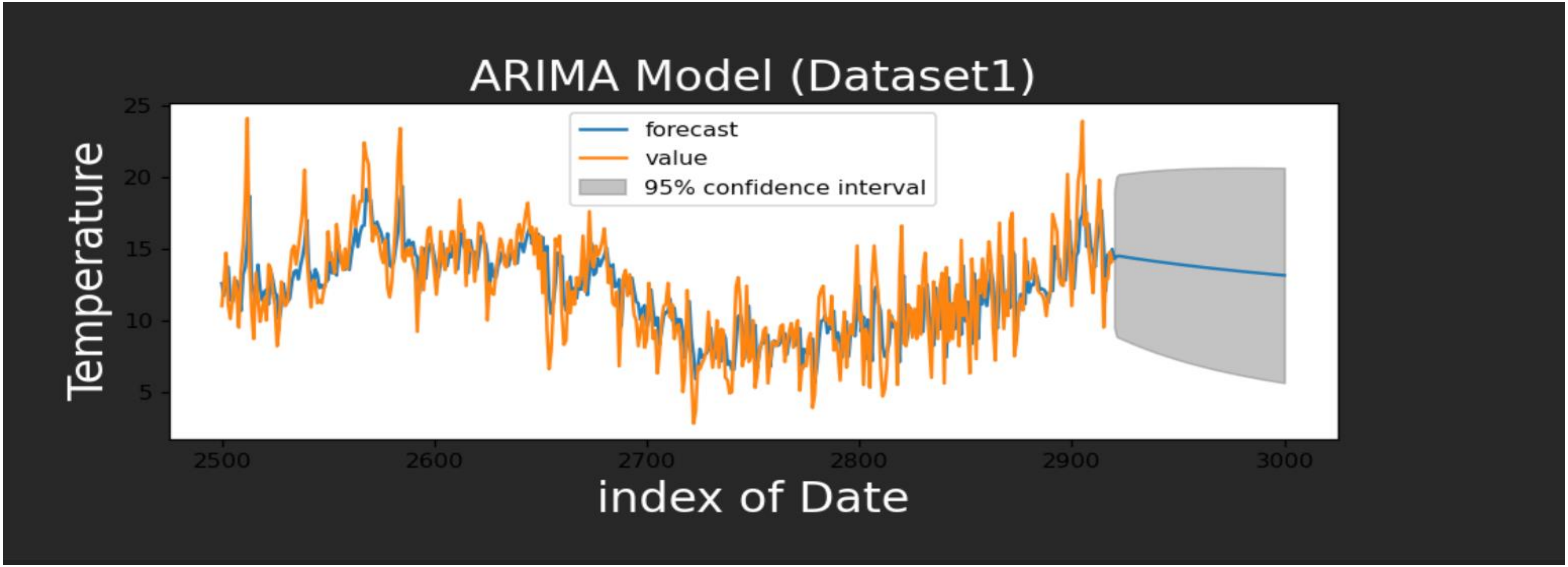
EUTS

Egyptian Universities Training Satellite

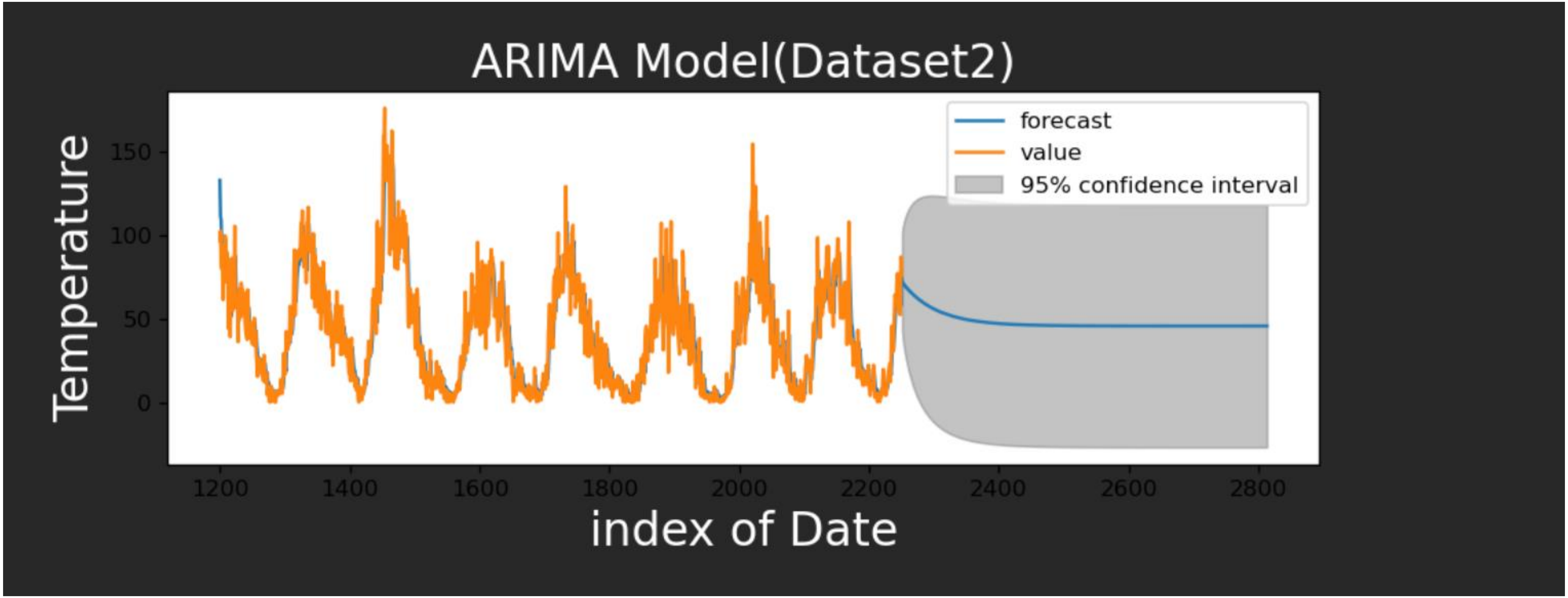
ARIMA Model

INTRODUCTION TO ARIMA

- An ARIMA model is a class of statistical models for analyzing and forecasting time series data.
- ARIMA is an that stands for Auto Regressive Integrated Moving Average.
- The statsmodels library provides the capability to fit an ARIMA model.
- An ARIMA model can be created using the statsmodels library as follows:
  - Define the model by calling ARIMA() and passing in the p, d, and q parameters.
  - The model is prepared on the training data by calling the fit() function.
  - Predictions can be made by calling the predict() function and specifying the index of the time or times to be predicted
- The parameters of the ARIMA model are defined as follows:
  - p: The number of lag observations included in the model, also called the lag order.
  - d: The number of times that the raw observations are differenced, also called the degree of differencing.
  - q: The size of the moving average window, also called the order of moving average.



Result “Mean Square Error:4.01



Result "Mean Square Error:62.04

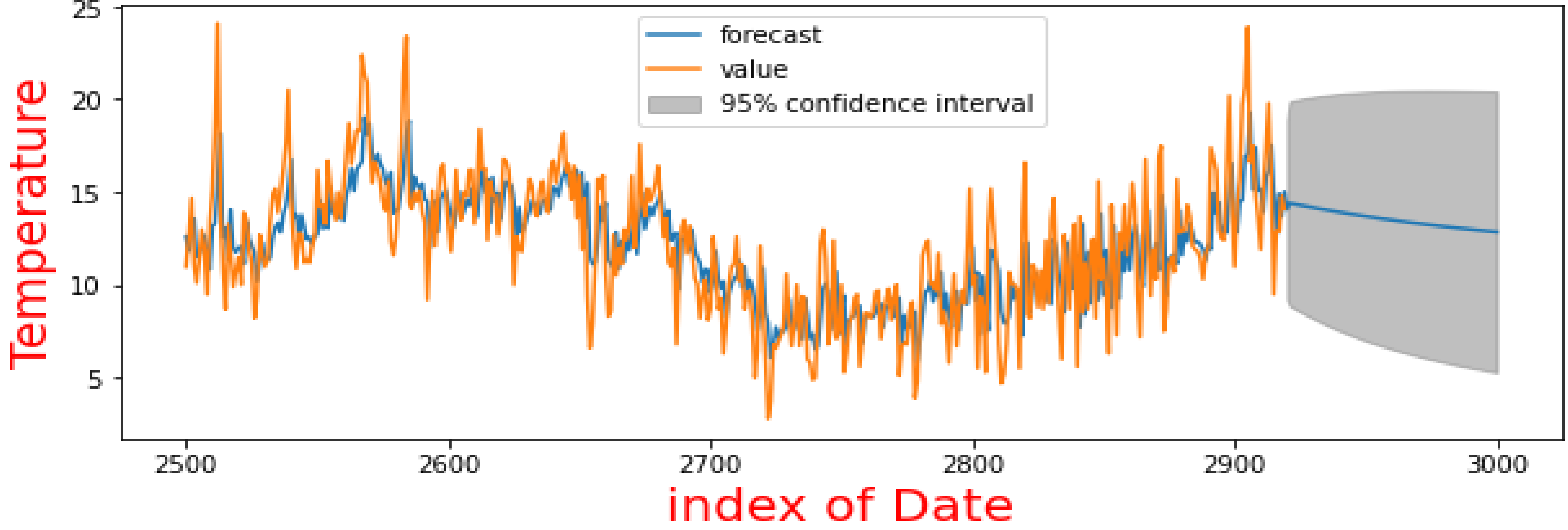
ARMA Model

INTRODUCTION TO ARMA

- ARMA stands for Auto Regressive Moving Average.
- An ARMA model is a class of statistical models used to describe weakly stationary time series.
- We use ACF & PACF plots to check parameters and see if there's a need to use differencing.
- An ARMA model can be created using the statsmodels library as follows:
  - Define the model by calling ARMA() and passing in the p, d, and q parameters.
  - The model is prepared on the training data by calling the fit() function.
  - Predictions can be made by calling the predict() function and specifying the index of the time or times to be predicted
- The parameters of the ARMA model are defined as follows:
  - p: The number of lag observations included in the model, also called the autoregressive order.
  - d: The number of times that the raw observations are differenced, also called the degree of differencing.
  - q: The size of the moving average window, also called the order of moving average.

Egyptian Universities Training Satellite

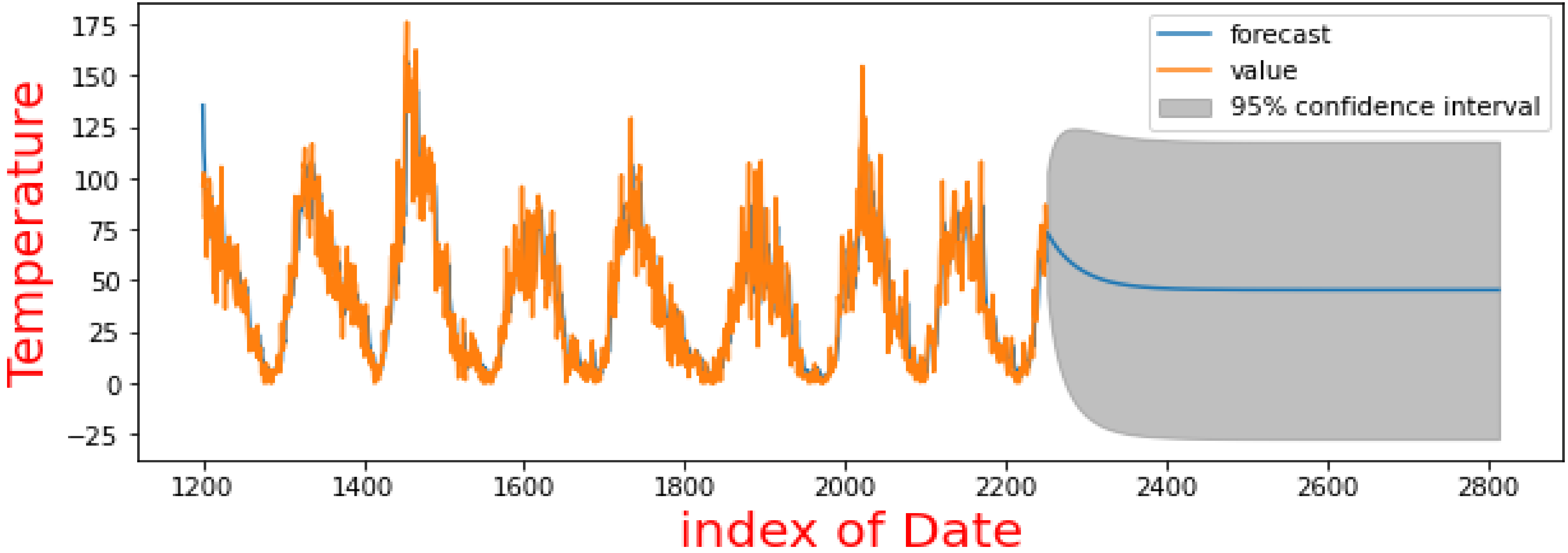
ARMA Model (Dataset1)



Result “Mean Square Error:3.987

EgSA
وكالة الفضاء المصرية
Egyptian Space Agency

ARMA Model(Dataset2)



Result “Mean Square Error:62.247

SES Model

Egyptian Universities Training Satellite

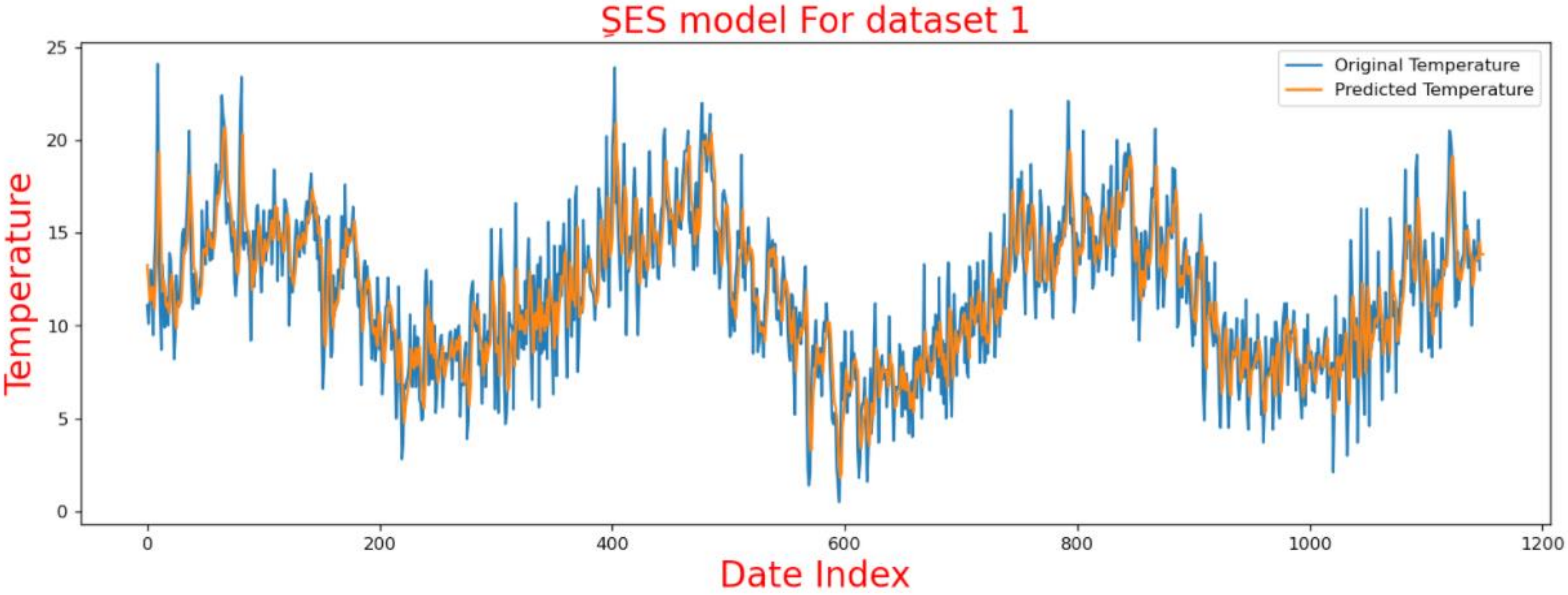
INTRODUCTION TO SES

- Exponential Smoothing is an elementary and pragmatic technique used for forecasting where the forecast is made through the exponentially weighted average of prior observations.
- It analyzes data from a specific period of time via providing more importance to recent data and less importance to former data
- This method produces “smoothed data”, the data that has a noise removed, and allows trends and patterns to be more clearly visible.

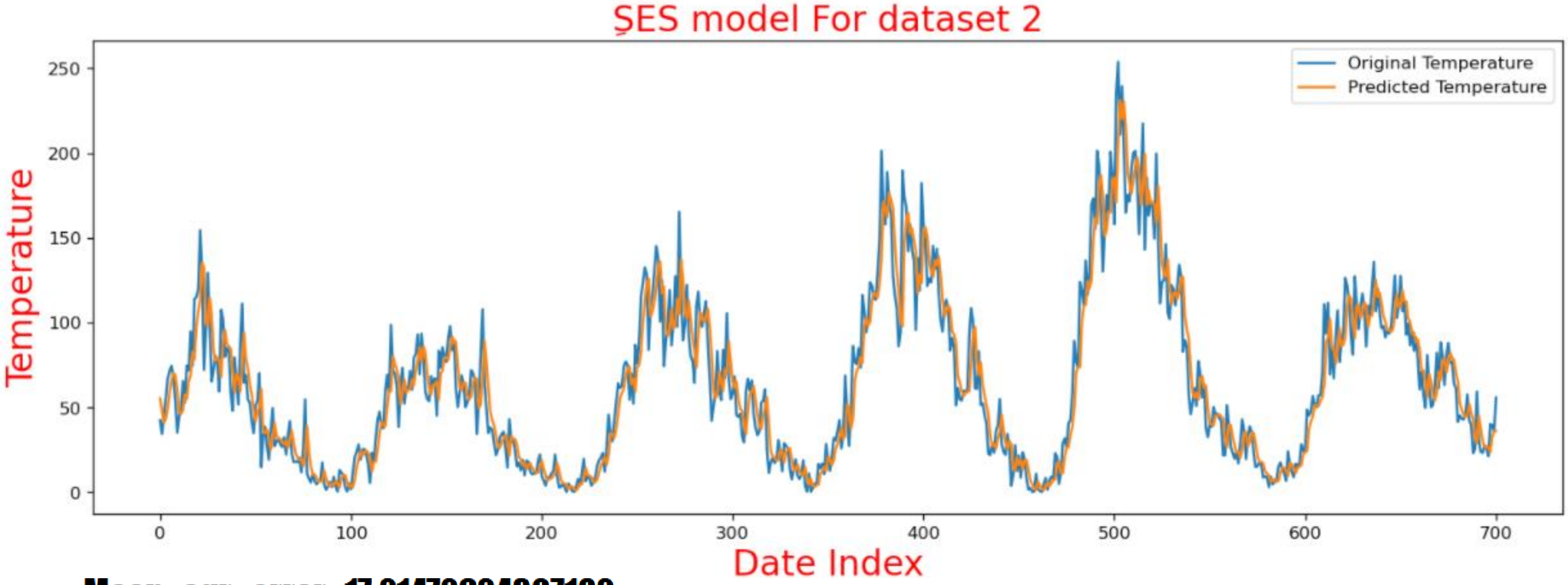
وكالة الفضاء المصرية

Egyptian Space Agency

- **Idea:-**Forecast future values using a weighted average of all previous values in the series.
- **Uses:-**Forecast a series with no trend and no seasonality
- **Types:-** 1-Simple exponential Smoothing (For series with no trend or seasonality) Good for our case!
- 2-Holt's Exponential Smoothing (For series with trend but no seasonality)
- 3-Winter's Exponential Smoothing (For series with trend and seasonality)
- **Key concepts:-**Smoothing Constant
- **Advantages:** Simple, Popular and adaptive0



Mean_sqr_error=2.16876874275847



Mean_sqr_error=17.21472624307182

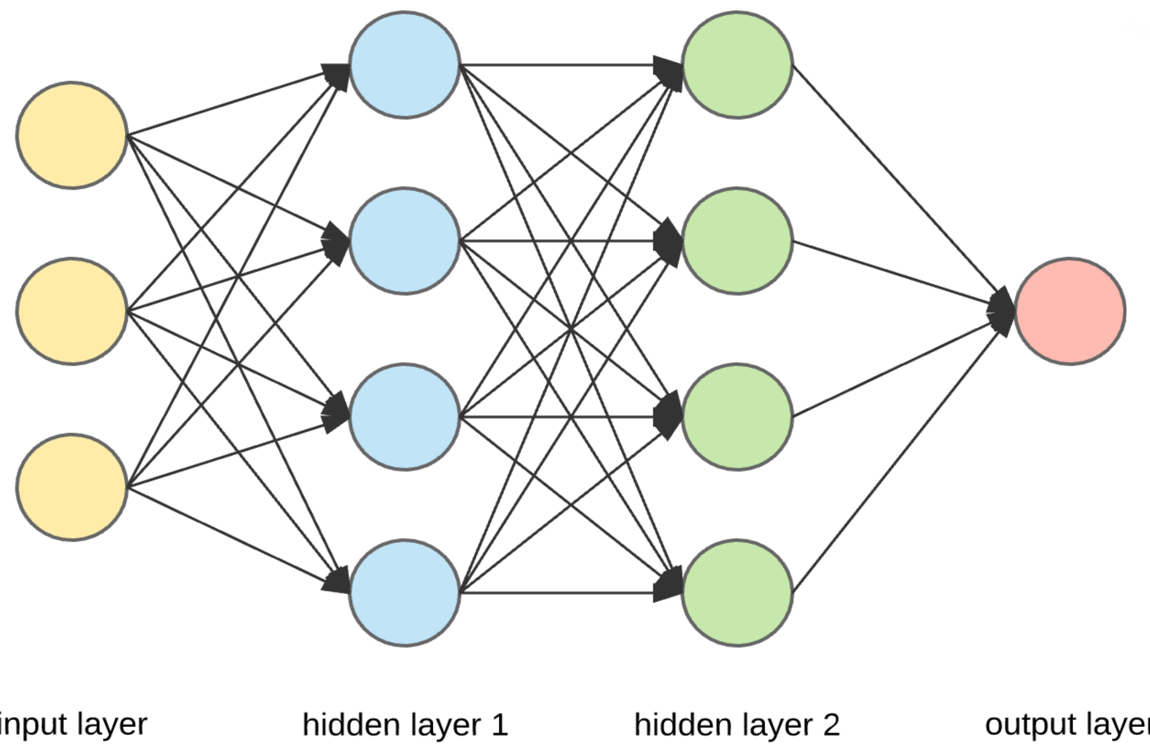
LSTM Model

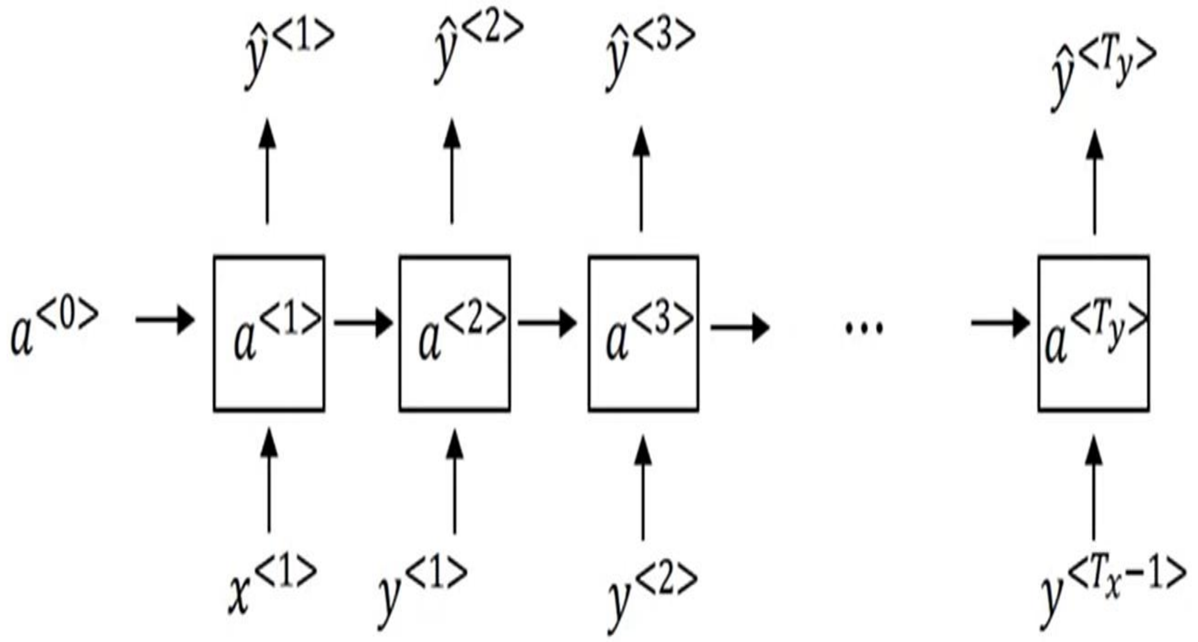
Basic Long Short Term Memory (LSTM)

Model type	application	The aim of impliment
LSTM is a type of deep learning model that is mostly used for analysis of sequential data (time series data prediction).	There are different application areas that are used: Language model, neural machine translation, music generation, time series prediction, financial prediction, etc.	The aim of this implementation is to help to learn structure of basic LSTM (LSTM cell forward, LSTM cell backward, etc..).

Standard NN

RNN

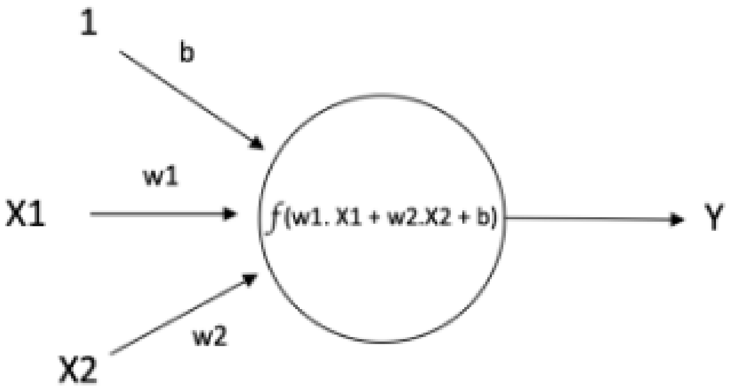




Egyptian Universities Training Satellite

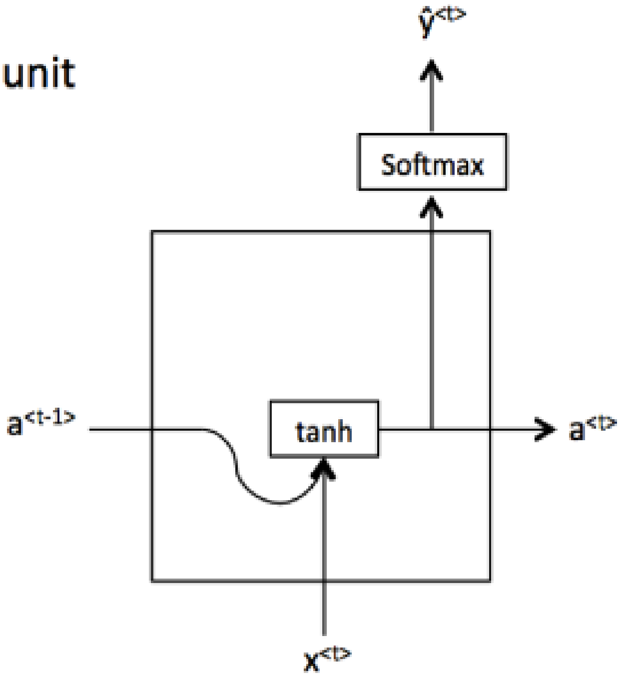
Types of NN

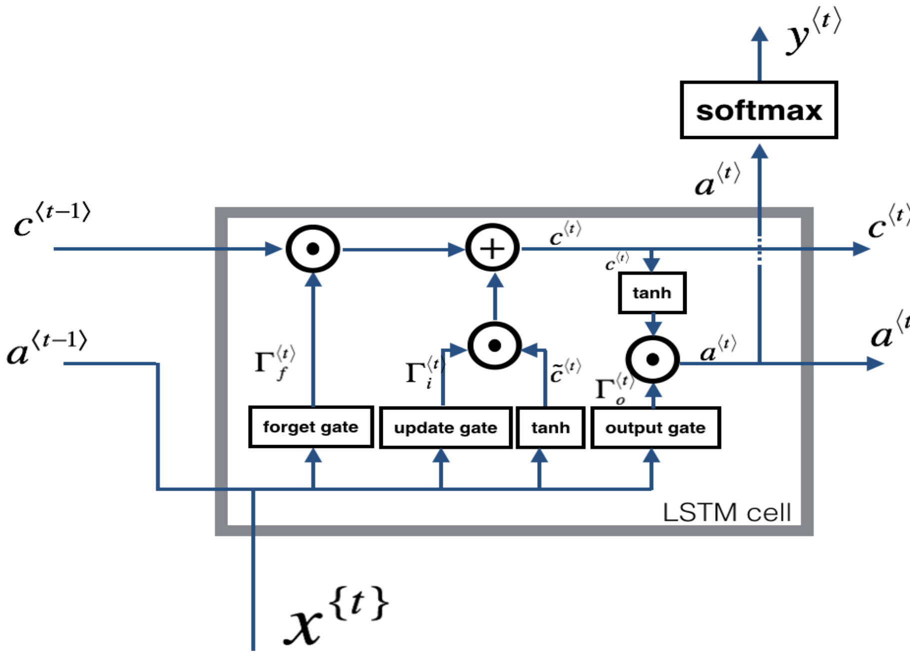
Standard NN | RNN | RNN with LSTM unit



$$Y = f(w1.\ X1 + w2.X2 + b)$$

RNN unit





LSTM Cell

Type of RNN units

$$\Gamma_f^{\langle t\rangle} = \sigma(W_f[a^{\langle t-1\rangle}, x^{\langle t\rangle}] + b_f)$$

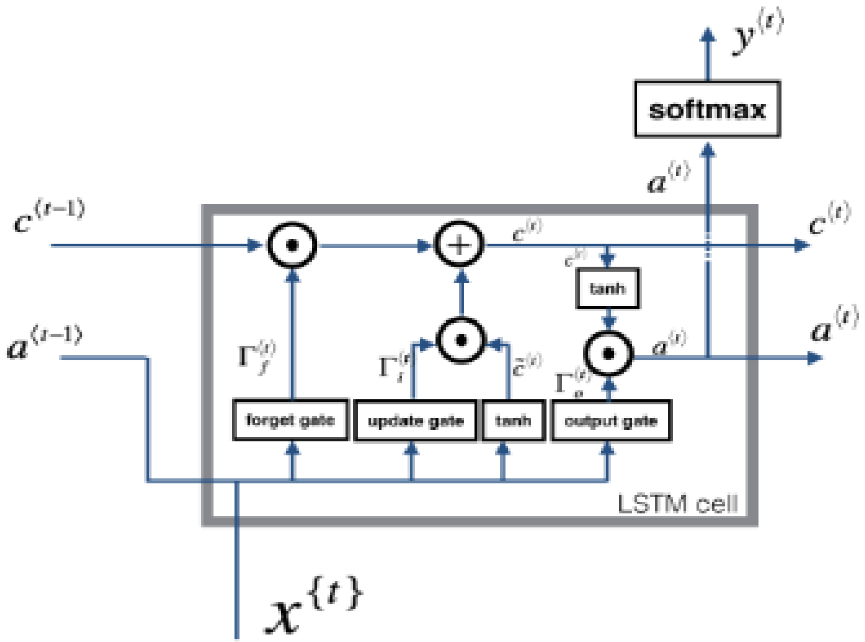
$$\Gamma_u^{\langle t\rangle} = \sigma(W_u[a^{\langle t-1\rangle}, x^{\langle t\rangle}] + b_u)$$

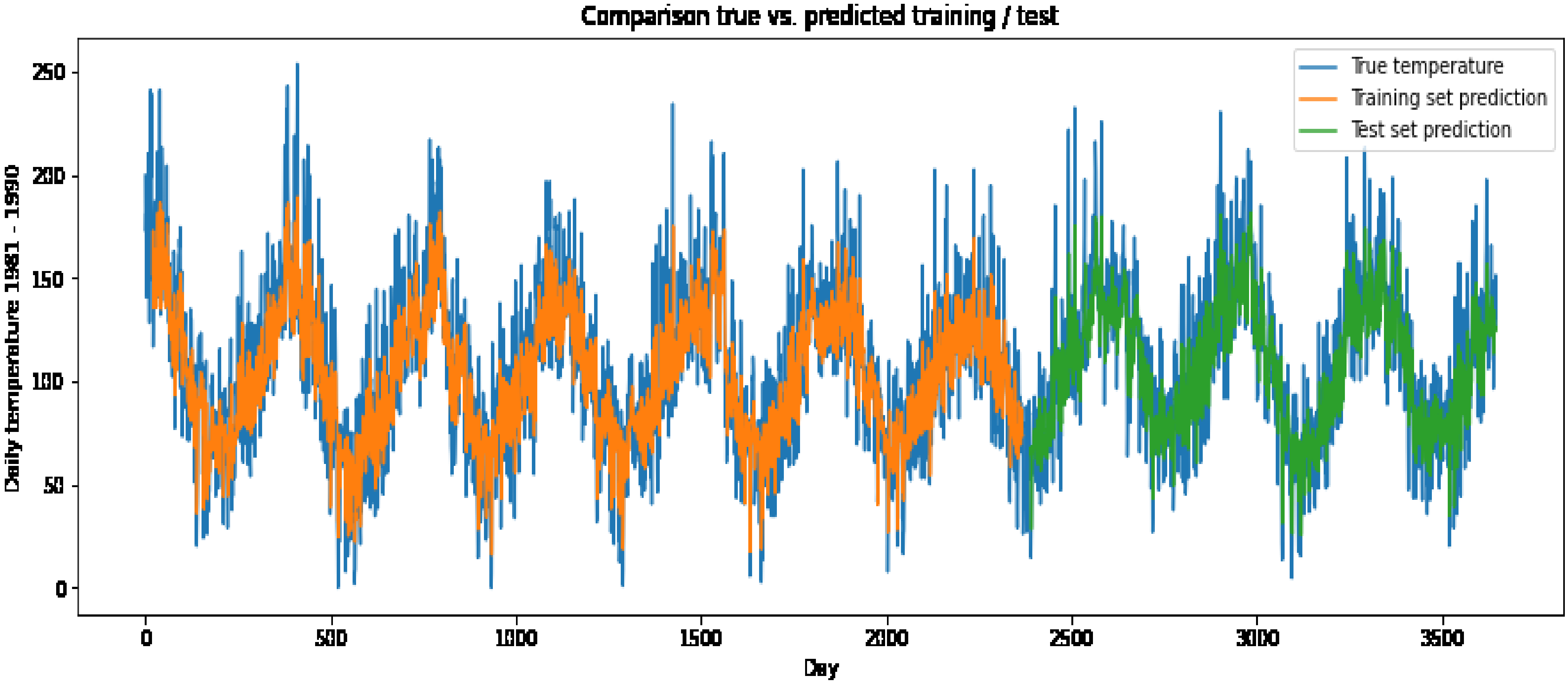
$$\tilde{c}^{\langle t\rangle} = \tanh(W_C[a^{\langle t-1\rangle}, x^{\langle t\rangle}] + b_C)$$

$$c^{\langle t\rangle} = \Gamma_f^{\langle t\rangle} \circ c^{\langle t-1\rangle} + \Gamma_u^{\langle t\rangle} \circ \tilde{c}^{\langle t\rangle}$$

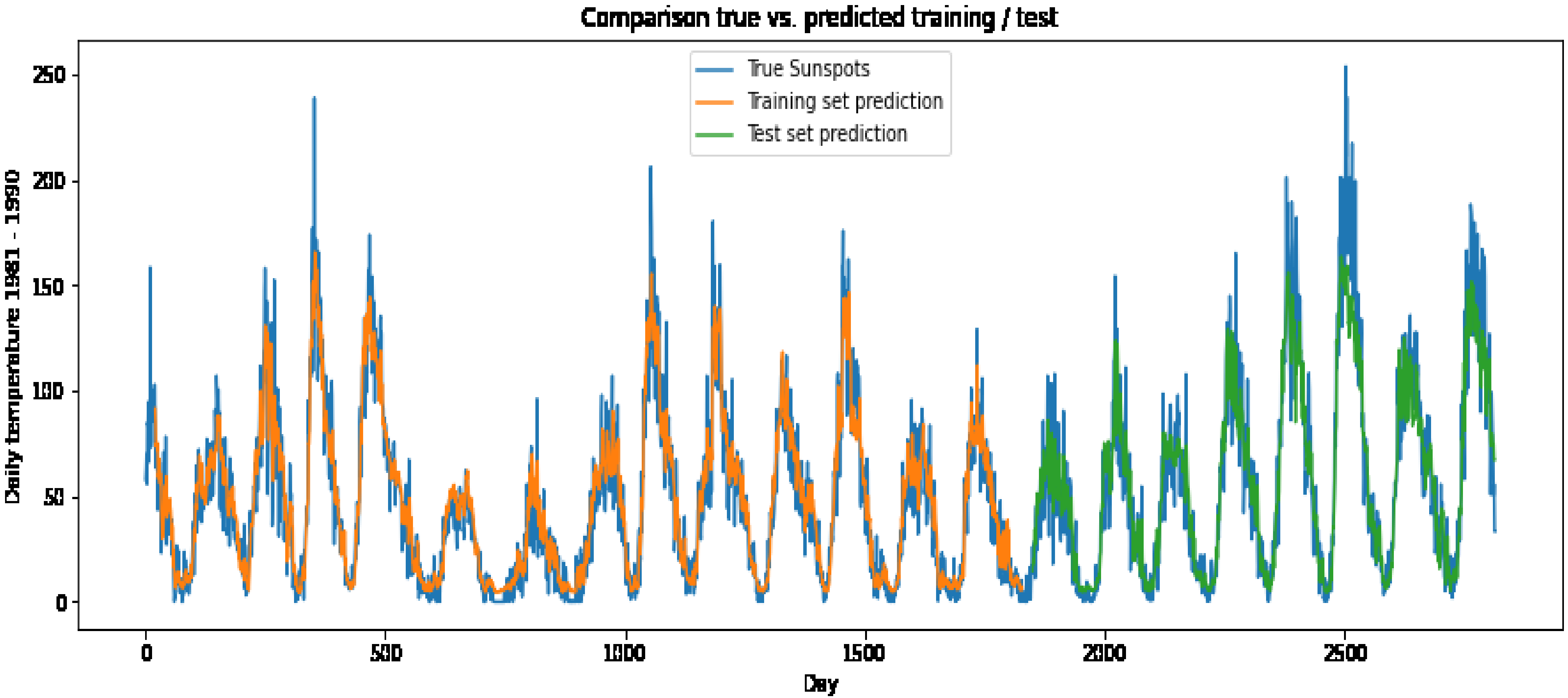
$$\Gamma_o^{\langle t\rangle} = \sigma(W_o[a^{\langle t-1\rangle}, x^{\langle t\rangle}] + b_o)$$

$$a^{\langle t\rangle} = \Gamma_o^{\langle t\rangle} \circ \tanh(c^{\langle t\rangle})$$





Comparison true vs. predicted training / test
True temperature
Training set prediction
Test set prediction
Daily temperature 1981 - 1990
Day
0
50
100
150
200
250
500
1000
1500
2000
2500
3000
3500


Comparison true vs. predicted training / test
True Sunspots
Training set prediction
Test set prediction
Daily temperature 1981 - 1990
Day
0
50
100
150
200
250
0
500
1000
1500
2000
2500

XGBOOST Model

Description

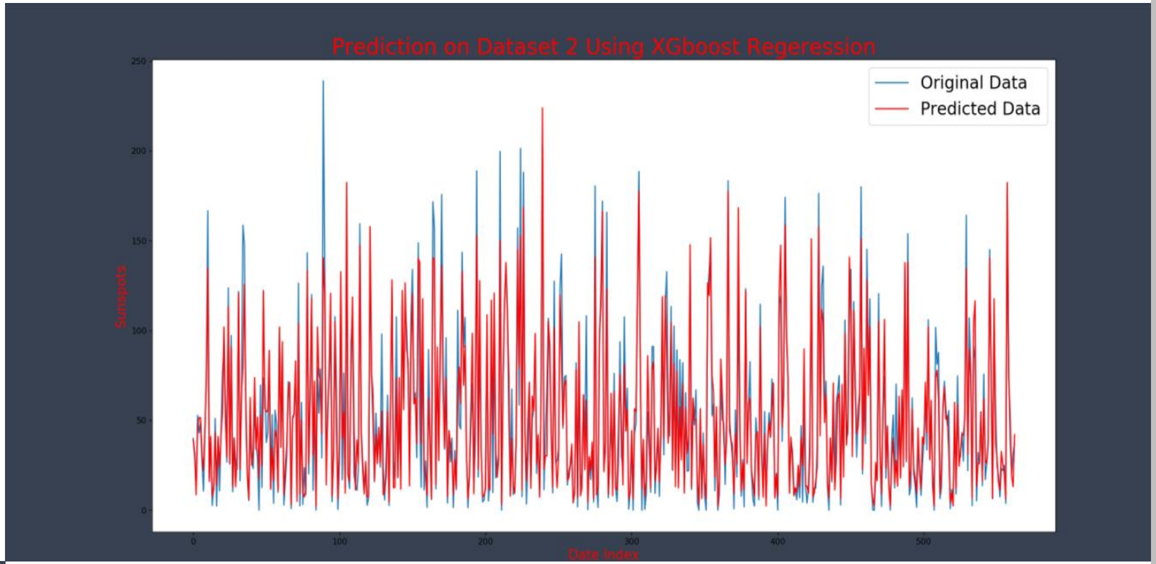
- XGBoost is an efficient implementation of gradient boosting for classification and regression problems.
- XGBoost can also be used for time series forecasting, although it requires that the time series dataset be transformed into a supervised learning problem first.

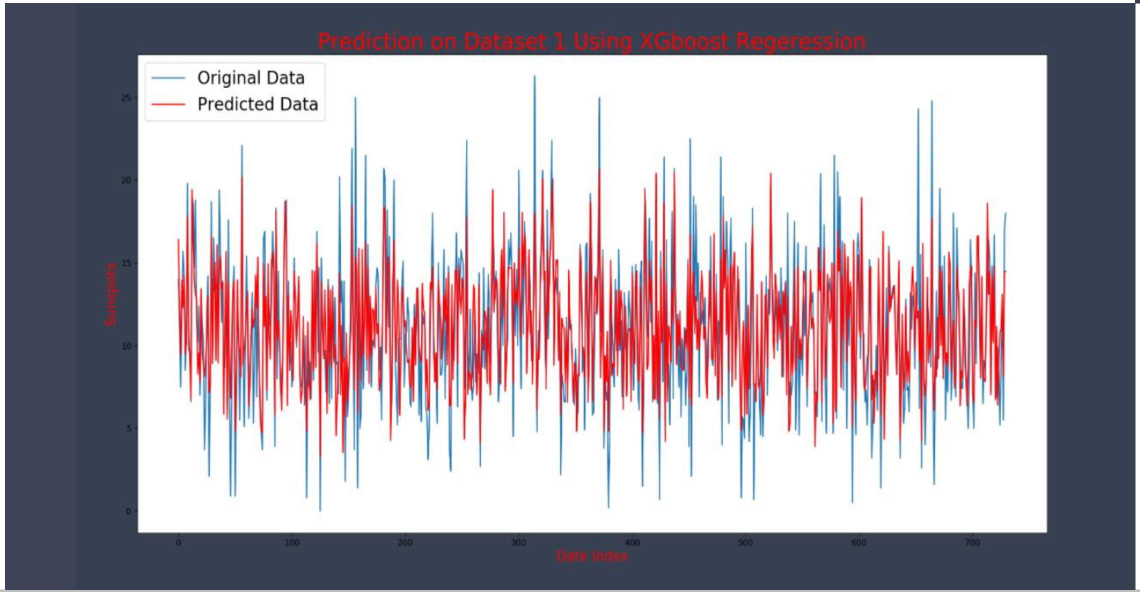
Egyptian Universities Training Satellite

implemented in project

- -First thig we Give Index Instead of Data to use XGBoost
- -then, we split data to train and test with size 0.8 to train and without random state
- - we try to find the best n_estimator through mean square error calculation then we select Best n_estimator for Dataset 1 is 100 and Dataset 2 is 70
- -after that we predict our data and here is the prediction result on plot:

Egyptian Universities Training Satellite





EgSA
وكالة الفضاء المصرية
Egyptian Space Agency

EUTS

Egyptian Universities Training Satellite

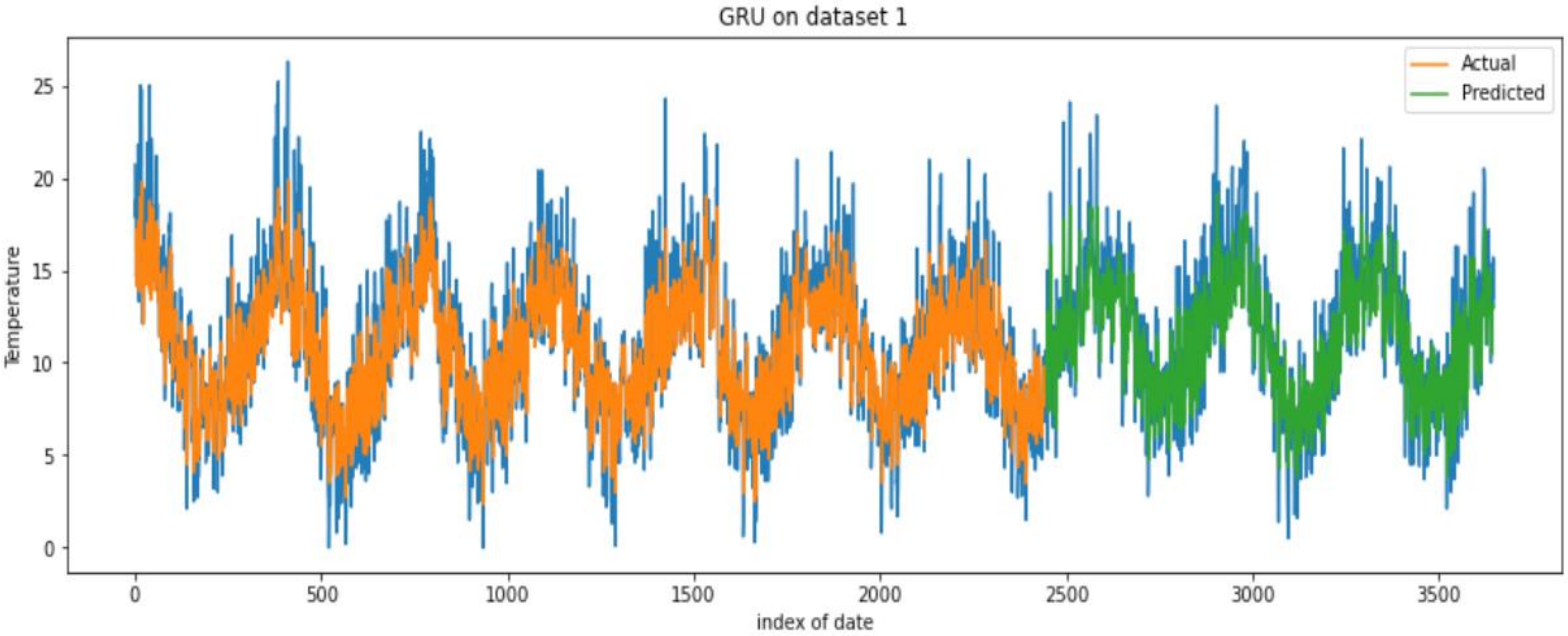
GRU Model

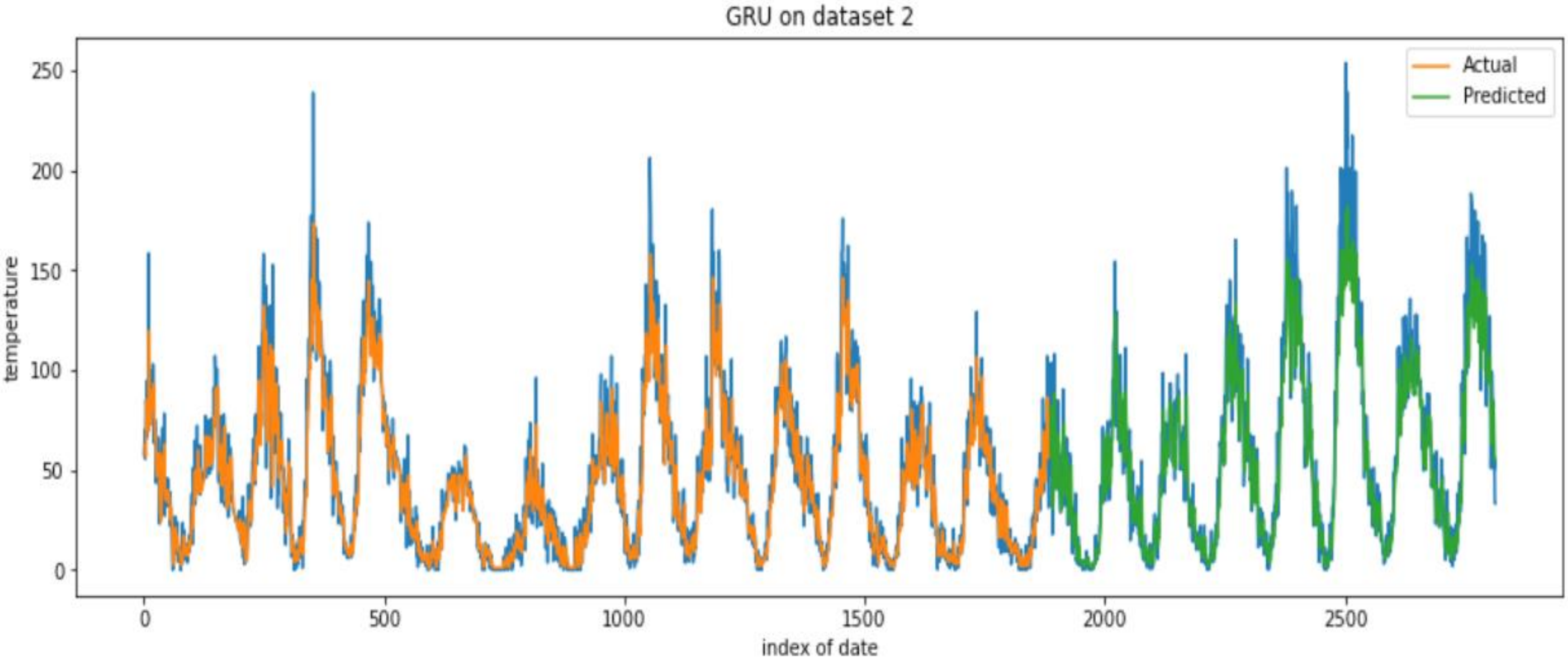
What is GRU

- Gated recurrent units (GRUs) are a gating mechanism in recurrent neural networks
- The GRU is like (LSTM) with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate.
- GRUs are improved version of standard recurrent neural network to solve the vanishing gradient problem of a standard RNN

How it works

- GRU uses update gate and reset gate. Basically, these are two vectors which decide what information should be passed to the output.
- they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction.
- The update gate helps the model to determine how much of the past information (from previous time steps) needs to be passed along to the future.
- The reset gate is used from the model to decide how much of the past information to forget.


GRU on dataset 1
Actual
Predicted
Temperature
index of date
0
5
10
15
20
25
0
500
1000
1500
2000
2500
3000
3500


GRU on dataset 2
Actual
Predicted
temperature
index of date
0
50
100
150
200
250
0
500
1000
1500
2000
2500

Algorithm Name	Mean Square Error on Dataset 1	Mean Square Error on Dataset 2
ARIMA	4.01	62.04
ARMA	**3.987**	**62.247**
Simple Exponential Smoothing "SES"	**2.16**	**17.21**
Long Short Term Memory 'LSTM'	23.15	18.49
XGBosst Regression	**1568.689**	**2091.9695**
Gated recurrent Unit "GRU"	2.56	15.42